Lab 9: Two Group Comparisons

Today’s Activities - Evaluating and interpreting differences across groups
- Effect sizes
  - Sex differences examples
  - Class example with data
  - Point biserial correlations
- Null Hypothesis testing
  - What does statistically “significant” mean?
- Comparing more than 2 groups
- Homework 9

Research Questions
- Lots of research questions center on whether there are differences between groups
  - Does an intervention work?
  - Do different groups show different levels of a construct of interest?
- Need methods to evaluate when differences are meaningful. Science provides guidelines to aid interpretation. (Still requires judgment!)
  - Effect sizes
  - Null Hypothesis testing
Basic Questions

• Are women and men basically different or the same when it comes to personality and other variables?
• Have these differences been exaggerated due to stereotypes?

Two Positions

• Minimalist Position – Sex differences are small and inconsequential.
  – Great variability within each sex
  – Most sex differences are actually quite small
  – Sex differences don’t matter for everyday life
• Maximalist – Sex differences exist and they matter.
  – Small effect sizes can have large practical consequences
  – Sex differences are comparable in size to many other effects that psychologists think are important
  – There is a range of variability in the sex differences
Thinking Meta-Analytically

- Lots of studies compare men and women on the average level of Variable X
- Goal of a Meta-Analysis: Collect and summarize the results of a body of literature.
- In a meta-analysis, researchers compute an effect size using the $d$ metric.
- $d = \text{Mean of Group 1} - \text{Mean of Group 2}$
  Pooled Standard Deviation

Cohen’s Rule of Thumb for $d$

- Small: Around .20 or -.20
- Medium: Around .50 or -.50
- Large: Around .80 and above or -.80 and below
- Comparisons:
  - Throwing distance: $d = 2.0$ (Men Longer)
  - Arm Strength: $d = 1.25$ (Men Stronger)
  - Height: $d = 1.75$ (Men Taller)

Sex Differences in Personality

Source: Feingold (1994)
Big Five Facet Differences

- Anxiety (N): $d = -.28$ (Women higher)
- Openness to Ideas (O): $d = .03$
- Assertiveness (E): $d = .50$ (Men higher)
- Activity (E): $d = .09$
- Trust (A): $d = -.25$ (Women higher)
- Tender-Mindedness (A): $d = -.97$ (Women higher)
- Order (C): $d = -.13$ (Women higher)

Other Differences

- Smiling: $d = -.60$ (Women smile more)
- Self-Reports of Attractiveness: $d = .24$ (Men higher)
- Actual Ratings of Attractiveness: $d = -.20$ (Men lower)
- Body Image/Satisfaction: $d = .58$ (Men higher)
- Aggression
  - Fantasy Measures: $d = .80$ (Men higher)
  - Peer Reports: $d = .63$ (Men higher)
  - Self-Reports: $d = .40$ (Men higher)
  - Worldwide Men commit about 90% of the homicides

Data Example –
Use file on class website
Are men and women different?

• Use class data for an example.
• Are there sex differences on the Agreeableness scale (agree)?

• Compute an effect size using the $d$ metric

\[ d = \frac{\text{Mean of Women} - \text{Mean of Men}}{\text{Pooled Standard Deviation}} \]

Calculate effect size for Agree

• IN SPSS:
  – Get the pooled SD
    • Analyze, Descriptives, Descriptives, agree
  – Get the mean scores for men and women
    • Data, Split file
    • Select ‘Organize output by groups’, select gender, OK
    • Descriptive Statistics, Descriptives, agree
  
  • Calculate by hand using the formula for $d$
    
    \[ d = \frac{\text{Mean of Group 1} - \text{Mean of Group 2}}{\text{Pooled Standard Deviation}} \]

Group differences – recast as point biserial correlations
**Point Biserial Correlations**

- Can also look at the same information in the context of the now-familiar correlation coefficient
- Use 0/1 variable to indicate group (Var: GenGroup)
  - E.g. Female group = 0, Male group = 1
- Run Correlations in SPSS (Agree and GenGroup)
- Interpretation is similar to a traditional correlation coefficient.
  - The point biserial correlation is positive when small values of X are associated with group=0 and large values of X are associated with group=1.

---

**Null Hypothesis Testing**

What do people mean by “significant”?

---

**Testing for Group Differences**

- Effect sizes describe the magnitude of difference between groups.
- Another way of interpreting the data is called null hypothesis testing
  - Set up a specific hypothesis
  - Set a decision criteria (specific threshold for rejecting or failing to reject the hypothesis)
  - Collect sample data. Evaluate hypothesis for “significance”
- Statistical significance depends on sample size; however, what really matters is often the effect size. (Is the difference meaningful?)
- Statistical Significance = Effect Size X Sample Size
Hypothesis testing with the t-test

- T-test is used when we’re interested in the mean difference between two sets of data.
- Does the data support rejecting the null hypothesis? If this support is obtained, then we can conclude that the mean of one group is significantly different from the mean of another group.
- What is the null hypothesis here?

Independent Samples T-Test – Hypothesis Testing

**Step 1:** Hypotheses and \( \alpha \)
- \( H_0: \mu_1 - \mu_2 = 0 \) (no difference between population means)
- \( H_1: \mu_1 - \mu_2 \neq 0 \) OR equivalently \( H_1: \mu_1 \neq \mu_2 \) (there is a mean difference between groups)
- \( \alpha = .05 \) and a two-tailed test will be used.

**Step 2:** Set decision criteria and locate critical regions.
- Example: To reject \( H_0 \), the obtained t-statistic must be < -2.101 or > +2.101.

**Step 3:** Collect sample data and calculate the t-statistic.

**Step 4:** Evaluate \( H_0 \).
- Example: The obtained t-statistic is greater than +2.101. (We reject \( H_0 \)).

T-test example in SPSS

- Example research hypothesis: Are there mean differences in hang over symptoms reported by men and women in PSY 395?
  - Identify:
    - Hypotheses
    - Decision criteria
- Calculate t in SPSS
  - Analyze, Compare Means, Independent samples t-test
  - Enter test variables (hang), and Grouping variables (gender)
  - OK
- Evaluate hypothesis and make decision (reject or fail to reject null hypothesis)
What if there are more than 2 groups?

More than 2 groups

- Example research hypotheses:
  - Is birth order (only child, 1st, 2nd, 3rd born) associated with different amounts of conscientiousness?
  - Is there a difference in the efficacy of different treatments for depression (individual psychotherapy, medication, group therapy)?
- Regardless of the number of groups, want to get a sense of the effect size.
  - How much do the groups differ?
- Often use theory to determine what comparisons are needed and meaningful.

Effect Sizes for Multiple Groups

- Imagine we are interested in studying how mood impacts performance on a task.
  - Create 3 experimental groups
    - Happy mood
    - Neutral mood
    - Sad mood
  - Examine differences of theoretical interest.
    - E.g., calculate an effect size comparing performance in the happy group with performance in the sad group.
Hypothesis testing with more than 2 groups

• Analysis of Variance (ANOVA)
  – Used to test for mean differences in dependent variable between groups/levels
  – Can use when there are more than two groups (which is the same as more than two levels of your IV)
  – Indicates if there is a mean level difference between groups. (Post hoc tests for details)
  – ANOVA is significant if F-statistic is greater than the critical values of F

Repeated Measures

• When we want to compare mean differences in the DV for the same group of people measured at two or more times
  – Does psychotherapy decrease depression? (where the same people would complete a depression measure both before and after receiving therapy)
• Can calculate effect sizes or use hypothesis testing.
  – Effect size (Mean at time 1 compared to mean at time 2)
  – Repeated measures T-test (2 groups)
  – Repeated measures ANOVA (more than 2 levels of IV.)

What you need to turn in for HW#9

• Use the class data to evaluate whether there are sex differences in the rest of the Big Five.
  – State the null hypotheses and the alternative hypotheses. (2 pts)
  – Calculate the d-metric effect size of this difference. Provide an interpretation of this estimate using Cohen’s guidelines. (2 pts)
  – Perform the same comparisons using a point-biserial correlation. Provide the interpretation for this estimate. (2 pts)
  – Use SPSS to perform an independent samples t-test to compare the means of the rest of the Big Five for men and women. What conclusions did you come to and why? (2 pts)
What you need to turn in for HW#9

- Imagine a researcher is interested in whether self-esteem varies based on year in college. State the null and alternative hypotheses the researcher would want to test. Identify what analysis could best determine whether there are mean level differences in self-esteem across 1st, 2nd, 3rd, and 4th year college students. (2 pts)